

ANÁLISIS DEL ESTADO DE LA PRESERVACIÓN DIGITAL EN LOS REPOSITORIOS DE LAS UNIVERSIDADES MEXICANAS

DAVID LEIJA y MIQUEL TÉRMENS

RESUMEN: Se presenta un estudio sobre el estado actual de preservación digital en los repositorios de las universidades mexicanas. Para ello se usó la metodología de análisis NDSA Levels, a partir de la cual en mayo-julio 2015 se pasó una encuesta a los responsables informáticos de estos repositorios. Se presentan los principales resultados obtenidos.

Palabras clave: Preservación digital; repositorios institucionales; universidades mexicanas; evaluación.

Keywords: Digital preservation; institutional repositories; Mexican universities; evaluation.

I. INTRODUCCIÓN

Desde hace más de una década las instituciones de educación superior de México están realizando un gran esfuerzo para mejorar la disponibilidad de información digital relacionada con la investigación. Estas mejoras se están aplicando a las bibliotecas digitales, a los repositorios institucionales y a las políticas de acceso abierto, entre otros ámbitos. Un avance importante fue la publicación por el Consejo Nacional de Ciencia y Tecnología (CONACYT) en noviembre de 2014 de los lineamientos generales para la creación del repositorio nacional y

de repositorios institucionales en las instituciones de educación superior mexicanas (CONACYT, 2014).

El documento de la CONACYT ha tenido una gran repercusión y está sirviendo de revulsivo para todo el sistema de soporte y difusión de la investigación en México. Posteriormente ha sido complementado por actuaciones ya concretas, por ejemplo en los apartados de financiación y de normalización técnica de repositorios. En este momento de mejora y cambio es adecuado preguntarse por cuál es la situación de partida, tanto para detectar los problemas del presente como, más adelante, poder evaluar los efectos de las políticas y actuaciones de promoción que se están iniciando.

Los objetivos contemplados en los lineamientos de CONACYT son realmente muy ambiciosos, por lo que en este estudio nos hemos centrado en uno solo de ellos, concretamente el objetivo tercero, que marca «Establecer las políticas que regulen la seguridad, almacenamiento, sostenibilidad, así como la gestión y preservación de la información». La investigación que presentamos pretende analizar cuál es el estado actual de las políticas de preservación digital que están aplicando los repositorios institucionales de información científica de las instituciones de educación superior de México (IES-MX).

En total existen 847 instituciones IES-MX de carácter público, pero el número de las que tienen repositorio institucional es mucho menor, concretamente de 9 IES federales y 34 estatales. Este es pues el universo sobre el que se ha realizado el estudio.

2. METODOLOGÍA

En el año 2010 se fundó la *National Digital Stewardship Alliance* (NDSA) (<http://ndsa.org/>) en los Estados Unidos, un consorcio de instituciones comprometidas en la preservación de recursos digitales. Sus actividades se centran principalmente en la difusión de buenas prácticas y en la mejora de la formación de los profesionales. En esta línea, dentro de NDSA se creó un grupo de trabajo con el encargo de crear una metodología que permitiera a las instituciones testear de forma fácil el nivel de desarrollo alcanzado por sus soluciones de preservación digital. El grupo de trabajo estuvo formado, entre otros, por representantes de la National Archives and Records Administration (NARA), el Metropolitan New York Library Council (METRO), la Harvard Library y la Library of Congress. Las propuestas de este grupo se presentaron a lo largo del año 2012 en distintas reuniones científicas y en ellas, incorporaron numerosas aportaciones de expertos en preservación digital. El redactado final de las propuestas se presentó en la conferencia *Archiving 2013*, celebrada en Washington DC en abril de 2013 (Levels, 2013) (Phillips, Bailey, Goethal, & Owens, 2013).

El documento de propuestas, conocido como *NDSA Levels*, consiste en una única tabla formada por preguntas sobre 5 ámbitos: sistema de almacenamiento y ubicación geográfica de los ficheros, alteración de los ficheros e integridad de los datos, medidas de seguridad de la información, metadatos y formatos de los ficheros. Para cada uno de los 5 ámbitos se presentan tareas a realizar ordenadas en 4 niveles, desde el nivel más bajo de exigencia, o nivel primario (nivel 1) al nivel más alto, o nivel completo (nivel 4) de preservación digital.

El uso de la tabla *NDSA Levels* se contempla como una vía fácil para recopilar el estado de la preservación digital en una organización concreta. Desgraciadamente la metodología concreta que se use para la compilación de los datos no está concretada en la propia tabla y existen diversas alternativas posibles, cada una de las cuales con sus propios problemas de fiabilidad (Térmens & Leija, 2017).

Dada la dispersión geográfica del universo a estudiar en nuestro caso, se optó por recolectar los datos mediante una encuesta en línea dirigida a los responsables de los distintos repositorios. La encuesta se aplicó mediante el sistema Survey Monkey entre el 18 de mayo y el 8 de julio de 2015. Los destinatarios de la encuesta fueron los máximos responsables de la biblioteca o bien el responsable del sistema informático, según se decidió después de mantener un contacto telefónico previo con los mismos. El nivel de respuestas obtenido en la encuesta fue del 86%, según se desglosa en la tabla 1.

Tabla 1. Grado de participación en la encuesta

INSTITUCIONES DE EDUCACIÓN SUPERIOR	NÚMERO	CON RESPUESTA	PORCENTAJE DE PARTICIPACIÓN
Federales	9	8	89%
Estatales	34	28	82%
Total	43	36	86%

3. RESULTADOS

Con el fin de caracterizar los repositorios estudiados, se preguntó sobre el tipo de materiales depositados. La Tabla 2 muestra los resultados obtenidos. Como se puede leer, existe una gran variedad de materiales depositados, pero destaca la combinación entre materiales de producción propia, resultado de las tareas investigadoras, como tesis y artículos, junto con materiales de producción externa, usados como fuentes auxiliares a la investigación, como libros digitales y bases de datos. Esta combinación es extraña en bibliotecas

digitales de Europa, los Estados Unidos y otros países avanzados, en las que las dos tipologías de materiales se gestionan y se ofrecen a los usuarios de forma separada, pero es habitual en México y en otros países con menores capacidades tecnológicas.

Tabla 2. Materiales depositados en los repositorios

MATERIAL DEPOSITADO	RESPUESTAS POSITIVAS	PORCENTAJE
Tesis académicas	27	77%
Libros digitales	25	71%
Artículos de investigación	21	60%
Bases de datos	17	49%
Material educativo	16	46%
Videos	12	34%
Fotografías	9	26%
Audios	7	20%
Otro	5	14%

Pasando ya a los resultados propios de la tabla NDSA, se obtuvieron una gran cantidad de datos que presentamos a continuación de forma resumida.

3.1. ALMACENAMIENTO Y LOCALIZACIÓN GEOGRÁFICA

En este apartado se evaluaba el número de copias de seguridad que se realizan de los datos, la mayor o menor distribución geográfica de estas copias, así como aspectos como el control de la obsolescencia de los sistemas de almacenamiento.

Se obtuvo que el 25% de las instituciones tienen 2 copias de seguridad completas y separadas y el 3% tienen 3 copias. Por tanto el 72% de las instituciones o no realizan copias de seguridad o estas se encuentran almacenadas en el mismo servidor, con lo que su eficacia es baja ante un desastre grave.

También son bajos los niveles de documentación del tipo de soportes utilizados (25%), la unificación sistemática de soportes variados (CD-ROM, pendrive, disquete) en un único soporte (31%) o disponer de un plan de migración futura de soportes (28%).

3.2. NO ALTERACIÓN DE ARCHIVOS E INTEGRIDAD DE LOS DATOS

En este apartado se evaluaban las políticas que se aplicaban para evitar la alteración del contenido binario de los ficheros y de comprobación de su integridad.

Los datos muestran que en el momento de la ingesta de ficheros, el 39% de las instituciones comprobaban su integridad, pero que luego solamente el 19% continuaba realizando comprobaciones periódicas a lo largo del tiempo. Aún era menor (8%) el porcentaje de instituciones que guardaban registros de las comprobaciones realizadas que pudieran ser útiles en el caso de una auditoría o control extraordinario. En caso de corrupción de ficheros, solamente el 33% declaraba disponer de la capacidad para recuperar los datos originales.

Ante el peligro de los virus, únicamente el 33% de las organizaciones tenía medios técnicos para hacer una comprobación de antivirus en la totalidad de sus fondos digitales.

3.3. SEGURIDAD DE LA INFORMACIÓN

Este apartado básicamente comprobaba las políticas de control de acceso del personal propio a los sistemas informáticos.

Menos de la mitad de los repositorios (concretamente del 44%) podían identificar quién ha leído, escrito, movido o eliminado ficheros concretos. Pero aún eran menos (el 14%) los que podían realizar auditorías en los registros de eventos para identificar qué persona había realizado una determinada acción.

Por el contrario, el 48% de los repositorios no mantenían ningún tipo de restricción en cuanto a la manipulación de los ficheros ni limitaban qué empleados lo podían hacer.

3.4. METADATOS

Este apartado identificaba los metadatos asociados a los distintos ficheros y cómo estos ayudaban a su gestión.

El 69% de los repositorios declaraban que no disponían de un inventario completo de sus fondos y, por ejemplo, no eran capaces de identificar en qué soporte físico concreto estaba almacenado un determinado objeto digital.

Respecto al almacenamiento de metadatos, el 61% de las instituciones guardaban metadatos descriptivos o de tipo técnico, pero solamente el 36% también guardaba metadatos de preservación, tipo PREMIS. Aun así

únicamente el 6% documentaban en forma de metadatos las transformaciones o migraciones de formatos que habían aplicado a los ficheros.

3.5. FORMATOS DE FICHERO

En este apartado se identificaban las políticas que se aplicaban a los formatos usados en los ficheros.

El 56% de los responsables que contestaron el cuestionario afirmaron que participaban en la toma de decisiones que afectaban al uso de nuevos formatos en documentos creados dentro de su organización. Luego se reducía al 42% aquellos que disponían de inventario de todos los formatos almacenados en su repositorio, y aún bajaba más, al 28%, aquellos que disponían de capacidad para monitorear la posible obsolescencia de estos formatos.

4. DISCUSIÓN Y CONCLUSIONES

La tabla *NDSA Levels* está organizada en cuatro niveles, de menor a mayor cumplimiento de los requisitos de seguridad y preservación de la información. Supone que dentro de una gestión armónica una organización cualquiera primero cumplimentará los requisitos de los niveles inferiores antes de emprender requisitos más avanzados. Por tanto, en una situación ideal, una organización que cumple, por ejemplo, con el nivel 3, también está cumpliendo con los niveles inferiores 1 y 2. Desgraciadamente en el mundo real las cosas no siempre son así, y las organizaciones no siempre cumplen de forma escalonada con los diferentes niveles, sea por desconocimiento, por problemas técnicos, problemas económicos u otras razones.

En la tabla 3 se muestran los resultados resumen por niveles NDSA. A nuestro parecer en ella llaman la atención dos cosas: por un lado el bajo porcentaje de cumplimiento en cualquiera de los cuatro niveles y, por el otro, que este cumplimiento no muestre una distribución decreciente, con valores mayores en el nivel 1 (el más básico) y menores en el nivel 4 (el más difícil de alcanzar).

Tabla 3. Resumen de resultados según niveles NDSA

NIVEL NDSA	NÚMERO DE ELEMENTOS A CONTROLAR	PORCENTAJE DE CUMPLIMIENTO
Nivel 1	9	35%
Nivel 2	10	26%
Nivel 3	9	25%
Nivel 4	8	27%

En nuestro análisis de los repositorios mexicanos hemos encontrado estas incongruencias de forma habitual, con a veces tasas más elevadas de éxito en ítems de niveles superiores que en niveles inferiores. Es posible que alguna de estas situaciones sea el resultado de una mala comprensión de las preguntas y las respuestas a dar en el cuestionario utilizado, pero creemos que no deben ser de importancia, pues los datos recopilados fueron depurados antes de su análisis y se realizaron consultas telefónicas o por correo electrónico para clarificar determinadas respuestas no congruentes. Por tanto, debemos concluir que la mayoría de estas incongruencias son el fiel reflejo de las tareas de gestión de estos repositorios.

Todo ello nos indica que nos encontramos ante un sistema de repositorios institucionales con graves problemas de gestión y con no siempre una buena planificación y priorización de las políticas a aplicar en cada momento. Por último, debemos recordar que en esta investigación se han analizado las respuestas de 36 instituciones sobre las 43 que disponían de repositorio institucional, pero todas ellas no son más que una fracción del total de 847 instituciones públicas de educación superior. Se trata pues de un conjunto enorme de instituciones que en su mayoría ni siquiera ha empezado a andar en el campo de los repositorios institucionales y en la preservación de sus contenidos.

5. AGRADECIMIENTOS

Este trabajo se realizó en el marco de una tesis doctoral sobre la preservación digital en México y con el soporte del proyecto «El acceso abierto a la ciencia en España: evaluación de su impacto en el sistema de comunicación científica» (Plan Nacional, ref. CSO2014-52830-P). También ha recibido soporte del GRC *Cultura i continguts digitals: aspectes documentals, polítics i econòmics*.

6. BIBLIOGRAFÍA

- CONACyT (2014). *Lineamientos técnicos para el repositorio nacional y los repositorios institucionales*. México DF: CONACyT. <http://www.siicyt.gob.mx/index.php/normatividad/2-conacyt/4-conacyt/1499-lineamientos-tecnicos-para-el-repositorio-nacional-y-los-repositorios-institucionales/file>
- DRYDEN, J. (2011). Measuring Trust: Standards for Trusted Digital Repositories. *Journal of Archival Organization*, 9(2), 127-130. <http://dx.doi.org/10.1080/15332748.2011.590744>
- FAUNDEEN, J. (2014). Building Trust: NDSA Levels of Digital Preservation. *Digital Preservation*. http://www.digitalpreservation.gov/meetings/documents/ndiipp14/Faundeen_NDSALevels.pdf
- GOETHALS, A. (2013). An Example Self –Assessment Using the NDSA Levels of Digital Preservation. *CAIB Ipres*. <https://benchmarkdigitalpreservation.files.wordpress.com/2013/09/caiw2013goethals.pdf>
- ISO (2012). *ISO 16363:2012. Space data and information transfer systems – Audit and certification of trustworthy digital repositories*. Geneva: ISO
- LEVELS (2013). *Levels of digital preservation*. <http://ndsa.org/activities/levels-of-digital-preservation/>
- PHILLIPS, M., BAILEY, J., GOETHALS, A., & OWENS, T. (2013). The NDSA Levels of Digital Preservation: An Explanation and Uses. *Archiving*. http://digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf
- RODRÍGUEZ, T. (2013). *Documento de diagnóstico nacional de repositorios institucionales en la IES mexicanas mediante una muestra representativa*. México DF: REMERI. <http://www.remeri.org.mx/portal/img/documentos/Anexo3.pdf>
- TÉRMENS, M., & LEIJA, D. (2017). Auditoría de preservación digital con NDSA Levels. *El profesional de la información*, 26(3), 447-456. <https://recyt.fecyt.es/index.php/EPI/article/view/58730>
- VOUTSSÁS, J. (2012). Long-term digital information preservation: challenges in Latin America. *Aslib Proceedings*, 64(1), 83-96.